

Explainable deep learning for diagnosing acute lymphocytic leukemia using blood smear images

Sion Theodorus Syaron Darmawan, Nabila Husna Shabrina

Department of Computer Engineering, Faculty of Engineering and Informatics, Universitas Multimedia Nusantara, Tangerang, Indonesia

Article Info

Article history:

Received Jul 29, 2024

Revised Oct 18, 2024

Accepted Nov 19, 2024

Keywords:

Acute lymphocytic leukemia
Deep learning
EfficientNetV2B3
Explainable AI
Local interpretable model-agnostic explanations
Microscopic images

ABSTRACT

Acute lymphocytic leukemia (ALL) is a rapidly progressing blood cancer that affects the lymphocytes. The diagnosis of ALL typically entails the examination of blood smears under a microscope, processes that are both time-consuming and susceptible to errors. Deep learning (DL) approaches have shown significant promise in automating the classification of ALL from microscopic images. However, the lack of transparency in these models hinders their widespread adoption in clinical settings. This study addresses this challenge by employing fine-tuned EfficientNetV2B3, a DL model, in conjunction with local interpretable model-agnostic explanations (LIME), a technique for explainable artificial intelligence (XAI) technique, to classify microscopic images of ALL. The C-NMC 2019 dataset, which has been augmented to ensure class balance, was utilized for training and evaluation. The proposed approach achieved impressive results, with an average recall, F1-score, and accuracy of 0.9795 and precision of 0.9796. The use of LIME effectively highlights relevant areas for prediction, accurately corresponding to the cell characteristics. The integration of DL and XAI techniques enhances the interpretability of ALL classification models, potentially increasing their trustworthiness and adoption in clinical practice. This study aims to further the development of diagnostic tools that are both precise and transparent for ALL.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nabila Husna Shabrina

Department of Computer Engineering, Faculty of Engineering and Informatics

Universitas Multimedia Nusantara

St. Scientia Boulevard, Gading Serpong, Tangerang, Banten 15811, Indonesia

Email: nabila.husna@umn.ac.id

1. INTRODUCTION

Leukemia comprises a diverse group of blood cancers that result from the abnormal and uncontrolled growth of immature white blood cells [1]. Leukemia can be classified based on its progression—acute or chronic—and the specific cell lineage involved—myeloid or lymphoid cells. Acute lymphocytic leukemia (ALL), commonly referred to as acute lymphoblastic leukemia (ALL), is a major concern. ALL is characterized by its rapid onset and progression, arising from the early forms of lymphocytes, a type of white blood cells. The global incidence of ALL has increased by 30.81% [2]. ALL is most commonly diagnosed in children under the age of five and adults over the age of 50 years. Approximately 25% of pediatric cancers are caused by ALL [3]. In Indonesia, 3,434 new cases of ALL are estimated to occur in children [4]. The prognosis for ALL varies significantly with age; while the disease progresses swiftly, only approximately 30% of adults with ALL can be cured [5]. Diagnosing conditions at an early stage and providing prompt treatment are essential for enhancing the patients' quality of life.

ALL is typically diagnosed by examining white blood cells using a blood smear through a microscope. Morphological evaluation is a valuable test for confirming the diagnosis and risk stratification of ALL [6]. The current diagnostic approach predominantly relies on manual methods in which the doctor visually analyzes the blood smear under a microscope. Nonetheless, differentiating malignant ALL cells from normal cells under a microscope is difficult due to their morphological similarities. In addition, this traditional method is time tedious and error-prone. Therefore, an automated system for ALL screening is required to significantly enhance diagnostic results. Automated systems can offer greater accuracy and faster diagnosis than manual methods, aiding doctors in providing more efficient treatment. Additionally, these systems could play a crucial role in rural and underserved areas, where medical experts are scarce [7].

Artificial intelligence, particularly through deep learning (DL) and computer vision techniques, presents a promising solution to these challenges [8]. Numerous studies have explored the use of DL for classifying leukemia, especially ALL. Sampathila *et al.* [9] developed a customized DL model based on convolutional neural networks (CNN) that achieved remarkable performance, including specificity of 95.81%, a sensitivity of 95.91%, accuracy of 95.54%, precision of 96%, and F1-score of 95.43%. Batool and Byun [10] enhanced EfficientNetB3 with depthwise separable convolutions to improve efficiency and reduce computational load, achieving 96.81% accuracy for the multiclass leukemia dataset. Ullah *et al.* [11] designed an attention-based CNN model that improves VGG16 architecture with an efficient channel attention (ECA) module. This method focuses on the most relevant aspects of the input images, allowing the model to enhance the feature extraction and representation. The proposed model resulted in a mean test accuracy of 91.1% [11]. Cho *et al.* [12] adopted the vision transformer and CNN model on an ALL classification dataset consisting of 12,528 samples and achieved accuracies of 86.2% and 88.4%, respectively. Abhishek *et al.* [13] presented a dataset of 500 peripheral blood smear images consisting of acute myeloid leukemia, ALL, and normal samples. They utilized ResNet50 combined with a support vector machine and attained an accuracy of 95% in their study. Shafi *et al.* [14] implemented CNN to differentiate three types of ALL utilizing peripheral blood smear microscopic images and achieved an accuracy of 89%.

Despite numerous evaluations of various DL model architectures in previous studies, there is an ongoing need to assess model interpretability and feasibility in clinical settings [15], [16]. Models employing neural networks, particularly CNNs, are often characterized as black-box predictions. This is because of the opacity of the training and testing processes, as well as the complexity of the model itself, which makes it difficult to discern which features are most utilized in the classification process. Transparency is of paramount importance in the healthcare sector, and the ability to provide clear explanations of the predicted outcomes is critical. This study aims to bridge the existing gap by employing explainable artificial intelligence (XAI) techniques [17]-[19] to provide understanding into the prediction outcome of DL models.

This research aims to advance the current understanding of ALL identification by employing the cutting-edge EfficientNetV2B3 architecture [20], [21] in conjunction with a local interpretable model-agnostic explanation (LIME) technique [22]. The C-NMC2019 dataset [23] comprising 12,528 microscopic images of ALL collected from 118 subjects was used in this study. The present study builds a valuable contribution to the development of an automated screening system for ALL by implementing a hybrid approach that integrates a CNN-based model and XAI for both the classification and explanation of ALL microscopic images.

2. METHOD

2.1. Research workflow

Figure 1 outlines the implementation process of this study. The process began with the preprocessing of the C-NMC2019 dataset, which involved several steps. Initially, the images were resized to a uniform dimension to match the optimal input requirements of the model. The dataset was partitioned into training and test sets. Subsequently, data augmentation techniques are employed to equalize the number of images in each class. Following the preprocessing stage, EfficientNetV2B3 architecture was employed for model training. Once the model training is complete, the trained model undergoes an evaluation process using the F1-score, precision, accuracy, and recall. To gain insights of the model's prediction-making process, the LIME method is employed by visualizing the highlighted regions to ensure that the predictions align with the expected cell characteristics.

2.2. C-NMC2019 dataset

This study employed the C-NMC2019 dataset, which originally comprised 15,135 images of ALL obtained from 118 individuals. The images, formed of 450×450 pixels and a black background, was sourced from publicly available datasets [23]. The dataset comprises images of cells segmented from microscopic images that closely resemble real-world conditions, including the staining noise and illumination errors. Expert classification further categorizes each dataset into two classes: ALL (leukemia cells) and HEM

(healthy and non-cancerous cells). For this research, only images from the training and validation folders were used, as these datasets included ground-truth labels, whereas the test data lacked such labels and could not be utilized. This study utilized 12,528 images, including 8,491 ALL and 4,027 HEM images. Sample images of the dataset are shown in Figures 2 and 3.

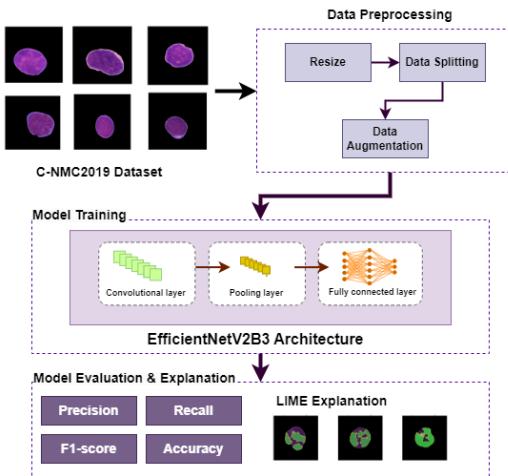


Figure 1. Research workflow

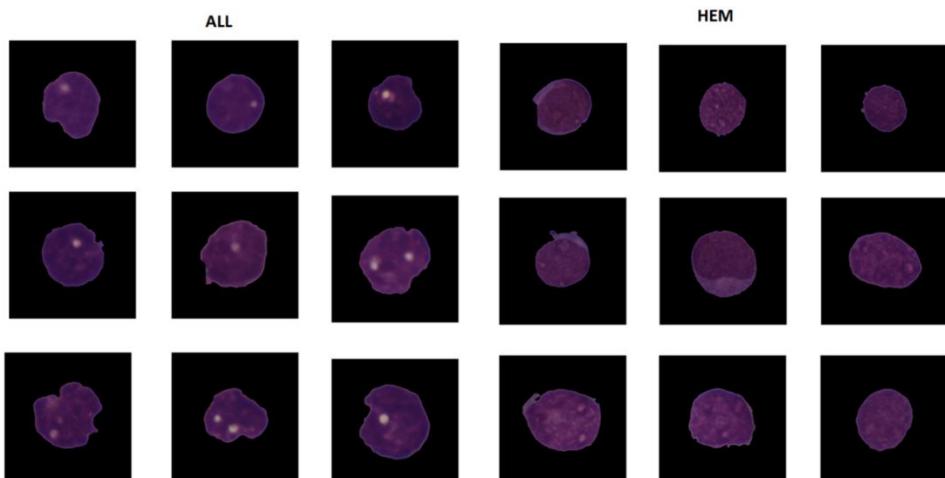


Figure 2. Sample for ALL class

Figure 3. Sample for HEM class

Typically, lymphocytes and lymphoblasts are round, with the nucleus occupying most of the cell area and the remaining portion being the cytoplasm. From the images of both classes, it is evident that most do not include the cytoplasm and change to a black background. This was likely because of the segmentation performed by the dataset creators, which resulted in the cytoplasm surrounding the nucleus of the lymphoblasts and lymphocyte cells being cut off. This is indicated by the images with overly rigid cuts and noncircular cell shapes. The segmentation process during the dataset creation is shown in Figure 4.

In the C-NMC2019 dataset, images of both ALL and acute lymphoblastic leukemia were provided. As shown in Figure 5, the white nucleoli, indicated by red arrows, are visible, and the nucleus surface appears smooth with minimal cytoplasm surrounding the cell, which is characteristic of lymphoblasts. Conversely, in lymphocyte cells, the nucleus surface was rough and uneven, the cytoplasm was relatively large compared to lymphoblasts (highlighted by green arrows), and no nucleoli or white dots were present. In the ALL class, which contains lymphoblast cells, some images display nucleoli very clearly, whereas in others, the nucleoli are either difficult to discern or are not visible at all. Similarly, in the HEM class images, some exhibited a rough surface on the nucleus, whereas in others, this rough texture was not as apparent. In

addition, in both classes, the cytoplasmic region is often unclear and is likely to be truncated during segmentation. This mixture of images, in which not all show the distinct characteristics of each cell type, mirrors real-life conditions in which microscopic images may not consistently display these features.

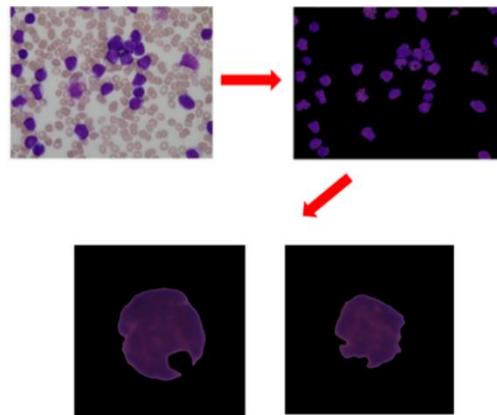


Figure 4. Segmentation process during dataset creation

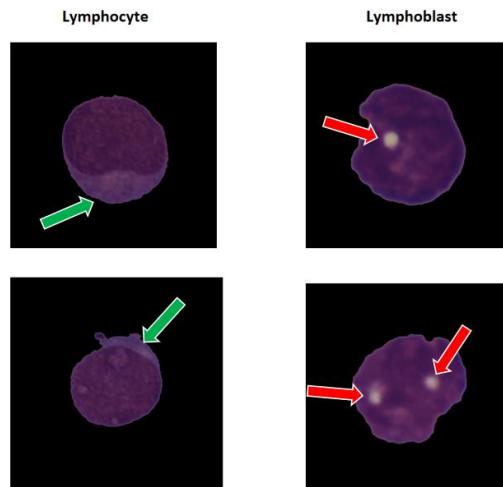


Figure 5. Example of lymphocyte and lymphoblast

2.3. Dataset preprocessing

The dataset was subsequently divided into train, validation, and test set with a ratio of 70:20:10. Owing to the uneven distribution of training data between the two classes, oversampling was applied through augmentation using ImageDataGenerator to ensure that both classes had a total of 10,000 images. The specific augmentation techniques employed were rotation with a range of 20°, vertical flip, and horizontal flip, which were based on previous research [9], [10]. These techniques were selected because they only involve rotating and flipping the image, thus minimizing the risk of the cell images being cut off owing to image shifting. In addition, the image size was adjusted to 300×300 pixels to accommodate the expected input shape of the EfficientNetV2B3 model [24].

2.4. Classification system

The classification system used in this study was constructed using EfficientNetV2B3. This model was selected because it is a more recent version of the EfficientNet family [21]. The EfficientNetV2 model demonstrated exceptional accuracy on the ImageNet dataset. Furthermore, it has the advantage of being smaller than the other CNN models [24]. The architecture of EfficientNetV2B3 is illustrated in Figure 6.

Several modifications were made to optimize the hyperparameters for model training. Adam optimizer was utilized with a batch size of 32, a learning rate of 0.0001, and 60 epochs. The Adam optimizer was selected based on its reputation as a powerful and adaptable optimization strategy [8], [25]. A batch size

of 32 was selected to accommodate the limited computational resources. An epoch of 60 was selected to meet a balance between training stability and allow sufficient time for the model to learn the features from the images. To mitigate overfitting, the EarlyStopping technique was implemented to monitor the validation loss and stop training when there was no improvement for six successive epochs. The training process was implemented using the Google Colab-free version.

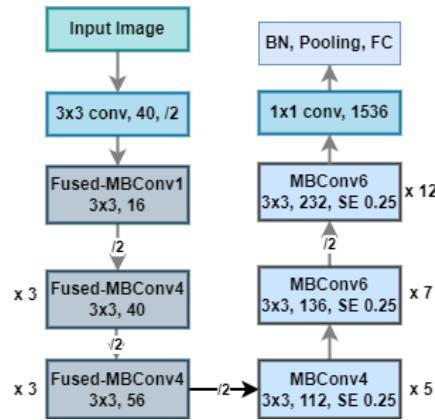


Figure 6. EfficientNetV2B3 architecture

2.5. Model evaluation and explanation

The key evaluation metrics used in this study were accuracy, F1 score, recall, and precision. In (1)-(4) provide the respective equations for these metrics, written in terms of true negatives (TN), true positives (TP), false positives (FP), and false negatives (FN). In addition, the learning curve was used to evaluate the performance during training by monitoring the accuracy and loss across multiple epochs [26]. This curve helps diagnose the model's proficiency in learning from the dataset.

$$Test\ Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

LIME is a technique used in the field of XAI. As the name suggests, LIME operates at the local level, focusing on a single instance, and is model-agnostic, meaning that it is compatible with any model. LIME perturbs the input data to produce multiple new synthetic datasets, which are variations of the original data. The new datasets were then assigned weights based on their resemblance to the original instance. LIME trains a model using synthetic data to make it easier to gain knowledge of the decision-making process from a more complex model. The goal was to develop a simpler model to replicate the behavior of the complex model, while capturing the overall features of the data. This allows LIME to identify the parts of the input data that the model relies on for its classification, thereby providing valuable information about the model's decision-making process. A mathematical explanation of the LIME is presented in (5). In general, the goal is to achieve a satisfactory approximation between the complex f model and simple g model.

$$explanation(x) = argmin g \in G L(f, g, \pi x) + \Omega(g) \quad (5)$$

where: x is the data to be explained, f is the complex model used, such as EfficientNet, g is one of the selected simple interpretable models for explanation, G is a collection of simple interpretable models, πx is proximity measure, and $\Omega(g)$ is regularization to ensure that the g model does not become excessively complex and remains interpretable.

3. RESULTS AND DISCUSSION

This study explored the performance of EfficientNetV2B3 in diagnosing ALL using the C-NMC2019 dataset. While previous research has focused on various CNN-based architectures for diagnosing ALL, they have often overlooked the crucial aspect of model explainability in their predictions. This study aims to bridge this gap by developing an explainable deep-learning model utilizing EfficientNetV2B3 along with LIME to enhance the transparency of the prediction results.

3.1. Model performance results

The model performance for training the C-NMC2019 dataset using EfficientNetV2B3 is listed in Table 1. As shown in Table 1, the average precision, recall, and F1 scores were 0.9796, 0.9795, and 0.9795, respectively. Furthermore, the model achieved an accuracy of 0.9795. We found that the addition of preprocessing through oversampling yielded consistent results for both ALL and HEM classes. The slight discrepancies in precision, recall, and F1-score among the classes indicated that oversampling was effective in enabling the model to learn from each class more effectively. The confusion matrix for these results is shown in Figure 7.

Table 1. Model performance results

Class	Precision	Recall	F1-score
ALL	0.9724	0.9870	0.9797
HEM	0.9868	0.9720	0.9794
Average	0.9796	0.9795	0.9795

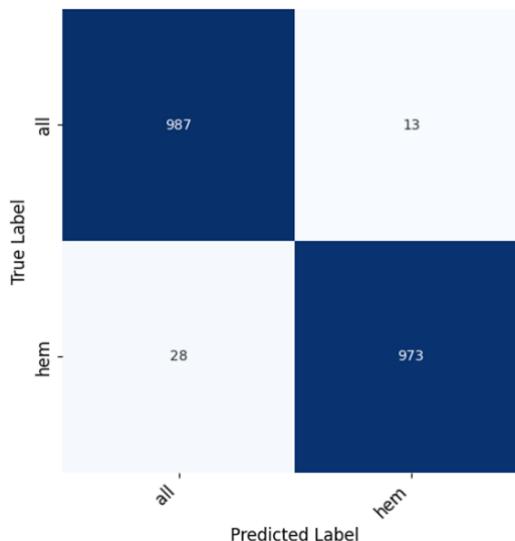


Figure 7. Confusion matrix

As shown in Figure 7, the confusion matrix reveals that the model makes predictions with high accuracy, with only 41 misclassified images. The majority of these misclassifications occurred in the HEM class (healthy lymphocyte cells), where the model incorrectly predicted the images as belonging to the ALL class. This misclassification trend can be attributed to the fact that the HEM class images did not differ significantly from ALL class images. Healthy lymphocyte cells in the HEM class typically have a larger cytoplasm compared to ALL cells, but the nucleus's color and texture in both classes are quite similar, making it challenging to differentiate them. Moreover, the original dataset contained approximately 7000 images for the ALL class and 3000 images for the HEM class. Consequently, the ALL class has more varied images despite previous data augmentation efforts. Although augmentation can modify existing images, it does not generate entirely new images, leading to less variety in the HEM class. As a result, the model tends to be more accurate in classifying images as ALL than as HEM.

The loss and accuracy curves for the model training process are shown in Figures 8 and 9, respectively. Figure 8 illustrates a gradual decrease in loss, approaching nearly zero, which signifies an effective error minimization during training. Both training and validation losses were very low, starting below 50% at the beginning of training and ending below 15% by the end. In Figure 9, the accuracy curve

consistently increased, indicating the model's successful feature learning from the data. This graph indicates that the model achieved a high accuracy in both the training and validation phases. The model achieved a maximum training accuracy of 98.7% and a validation accuracy of 97.5%. The small difference in accuracy between training and validation—only 1-2%—indicates that the model is neither overfitting nor underfitting. This performance surpasses that of previous research, in which the difference in accuracy between training and validation was as high as 5%. This improvement suggests that the data augmentation techniques implemented by the author effectively minimized overfitting in the model.

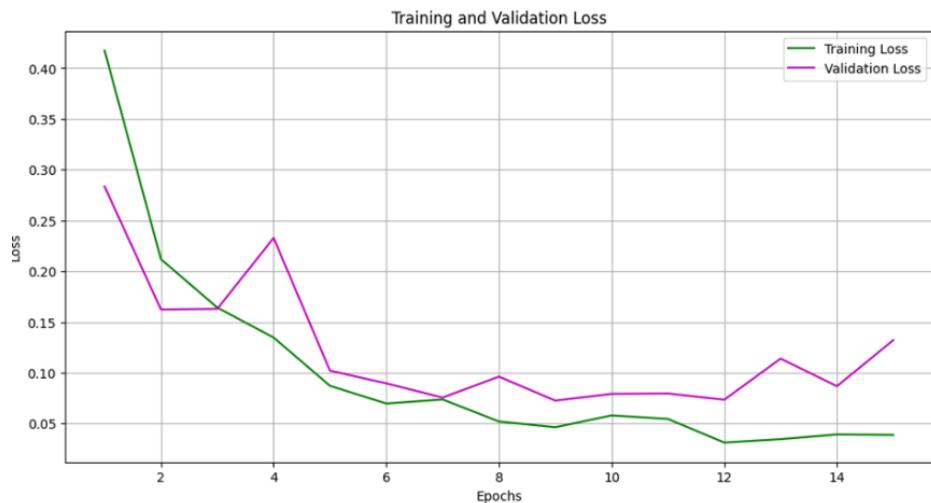


Figure 8. Training and validation loss curve



Figure 9. Training and validation accuracy curve

3.2. Model explainability

To test the LIME results, the author selected nine images from the ALL class (lymphoblast cells) and nine images from the HEM class (lymphocyte cells) were used. The images were selected to highlight their distinctive characteristics. Each image in Figure 10 and 11 (previously described in section 2) was analyzed using the LIME module, with the previously trained model as a reference. Figure 10 shows the LIME results for ALL classes. In Figure 10, the green-highlighted areas identified by the LIME model illustrate the features of the model used for its predictions. It is evident that the model is focused on the surface of the nucleus. For lymphoblast cells, the surface is smooth and homogeneous, and the model also identifies nucleoli, which appear as white dots within lymphoblast cells. Figure 11 shows the LIME results

for the HEM class. As can be seen in Figure 11, the green-highlighted areas indicate the parts of the image that the model used for classification, while the red-highlighted areas show the regions "discarded" by the LIME model, deemed less influential for classification. These images revealed that the model predominantly focused on the rough nuclear region of lymphocyte cells, as evidenced by the extensive green highlighting of the nuclear area.

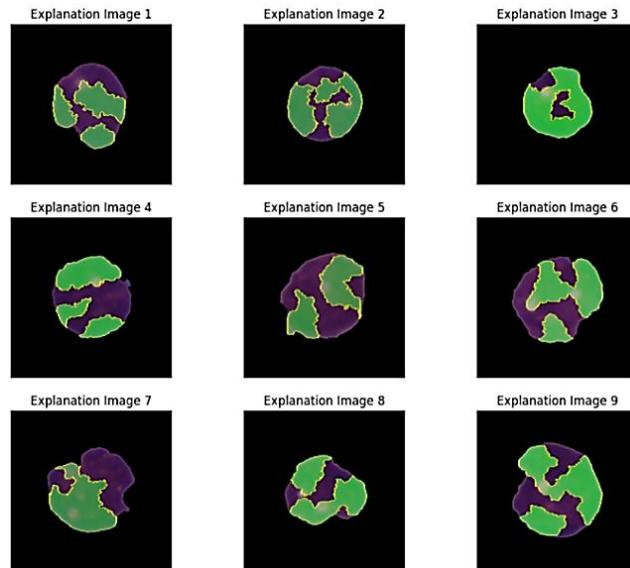


Figure 10. LIME explanation results on ALL class

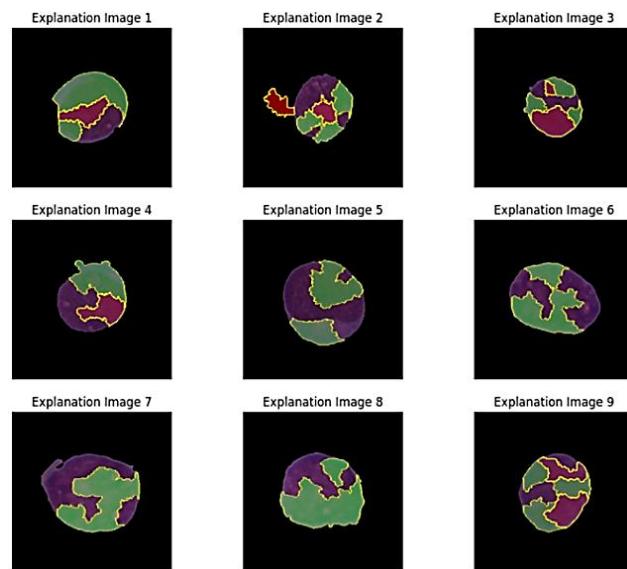


Figure 11. LIME explanation on HEM class

Figures 10 and 11 demonstrate that the model effectively identified relevant features for predictions corresponding to the distinctive characteristics of each cell type. Our findings indicate that the fine-tuned EfficientNetV2B3 model is not only accurate but also makes informed predictions by targeting appropriate cell features. As explained in the dataset section, the images did not include the cytoplasm, as they were removed during segmentation when the dataset was created. Typically, lymphocytes, and lymphoblasts are round, with the nucleus occupying most of the cell area and the remaining portion being the cytoplasm. Including the cytoplasm in the images could enhance the classification and explanation accuracy, because the

nucleus-to-cytoplasm ratio is a key factor in distinguishing between lymphocytes and lymphoblasts, especially when other distinguishing features are less visible.

3.3. Comparison with previous studies

Several studies have explored the use of CNN-based architectures for ALL diagnosis. This section compares our results with prior studies that also utilized the C-NMC 2019 dataset to emphasize progress and identify potential areas for enhancement. Cho *et al.* [12] adopted a CNN architecture and reported an accuracies of 86.2%. Similarly, Ullah *et al.* [11] utilized an attention-based CNN model to focus on the most relevant aspects of input images, achieving an accuracy of 91.1%. Our study, using the EfficientNetV2B3 model, demonstrated a significant improvement, with 11.75% and 6.85% higher accuracy, respectively. This performance boost likely stems from the augmentation techniques and EfficientNetV2B3 architecture's superior ability to extract features efficiently. Additionally, Batool and Byun [10] used the previous version of our model, EfficientNetV3, and achieved an accuracy of 96.81%. Although the difference in accuracy between their study and ours was small, they did not investigate which features of the images the model relied on to make its predictions. In contrast, our study not only achieves high accuracy but also prioritizes model explainability, offering insights into the decision-making process. However, while these results are promising, the model has not yet been tested in clinical environments. Further research is necessary to validate the clinical utility of our model through expert evaluation and real-world testing, to ensure its robustness and reliability for diagnostic use.

4. CONCLUSION

In this study, we developed an automated system capable of accurately and efficiently classifying ALL cells from microscopic images, while providing transparent explanations for its predictions. Utilizing EfficientNetV2B3 and LIME on the augmented C-NMC 2019 dataset. The results demonstrate that the proposed method achieves an average precision of 0.9796 with a recall, F1-score, and accuracy of 0.9795. The use of LIME effectively highlights relevant areas for prediction, corresponding accurately to the cell characteristics. Additionally, the small difference of approximately 1% between the training and validation accuracies indicates that the model does not suffer from overfitting. Our findings demonstrate that the proposed method not only achieves high performance, but also emphasizes model explainability, providing valuable insights into the decision-making process. However, while these results are promising, the model has not yet been tested in clinical environments, and its performance on a more diverse dataset needs further exploration. Future research could involve confirming the practical value of these results through expert assessment and real-world application, ensuring its reliability for ALL diagnosis, as well as expanding its application to other forms of leukemia.

ACKNOWLEDGEMENTS

The authors thank Universitas Multimedia Nusantara for the support provided during this study.

REFERENCES

- [1] A. Chennamadhavuni, V. Lyengar, K. R. M. Mukkamalla, and A. Shimanovsky, "Leukimia," National Library of Medicine. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK560490/>, (accessed: Jan. 01, 2024).
- [2] M. Yi, L. Zhou, A. Li, S. Luo, and K. Wu, "Global burden and trend of acute lymphoblastic leukemia from 1990 to 2017," *Aging*, Nov. 2020, doi: 10.1863/aging.103982.
- [3] D. Bhojwani, J. J. Yang, and C.-H. Pui, "Biology of Childhood Acute Lymphoblastic Leukemia," *Pediatric Clinics of North America*, vol. 62, no. 1, pp. 47–60, Feb. 2015, doi: 10.1016/j.pcl.2014.09.004.
- [4] D. Garniasih, S. Susanah, Y. Sribudiani, and D. Hilman, "The incidence and mortality of childhood acute lymphoblastic leukemia in Indonesia: A systematic review and meta-analysis," *PLoS ONE*, vol. 17, no. 6, p. e0269706, Jun. 2022, doi: 10.1371/journal.pone.0269706.
- [5] Y. Puckett and O. Chan, "Acute Lymphocytic Leukemia," National Library of Medicine, Aug. 26, 2023, [Online] Available: <https://www.ncbi.nlm.nih.gov/books/NBK459149/>, (accessed: Jan. 01, 2024).
- [6] T. Terwilliger and M. Abdul-Hay, "Acute lymphoblastic leukemia: a comprehensive review and 2017 update," *Blood Cancer Journal*, vol. 7, no. 6, pp. e577–e577, Jun. 2017, doi: 10.1038/bcj.2017.53.
- [7] S. Shafique and S. Tehsin, "Computer-Aided Diagnosis of Acute Lymphoblastic Leukaemia," *Computational and Mathematical Methods in Medicine*, vol. 2018, pp. 1–13, 2018, doi: 10.1155/2018/6125289.
- [8] A. E. Aby, S. Salaji, K. K. Anilkumar, and T. Rajan, "A review on leukemia detection and classification using artificial intelligence-based techniques," *Computers and Electrical Engineering*, vol. 118, p. 109446, Sep. 2024, doi: 10.1016/j.compeleceng.2024.109446.
- [9] N. Sampathila *et al.*, "Customized Deep Learning Classifier for Detection of Acute Lymphoblastic Leukemia Using Blood Smear Images," *Healthcare*, vol. 10, no. 10, p. 1812, Sep. 2022, doi: 10.3390/healthcare10101812.

[10] A. Batool and Y.-C. Byun, "Lightweight EfficientNetB3 Model Based on Depthwise Separable Convolutions for Enhancing Classification of Leukemia White Blood Cell Images," *IEEE Access*, vol. 11, pp. 37203–37215, 2023, doi: 10.1109/ACCESS.2023.3266511.

[11] M. Z. Ullah *et al.*, "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification," *Applied Sciences*, vol. 11, no. 22, p. 10662, Nov. 2021, doi: 10.3390/app112210662.

[12] P. Cho, S. Dash, A. Tsaris, and H.-J. Yoon, "Image transformers for classifying acute lymphoblastic leukemia," in *Medical Imaging 2022: Computer-Aided Diagnosis*, San Diego, United States: SPIE, Apr. 2022, p. 112. doi: 10.1117/12.2611496.

[13] A. Abhishek, R. K. Jha, R. Sinha, and K. Jha, "Automated classification of acute leukemia on a heterogeneous dataset using machine learning and deep learning techniques," *Biomedical Signal Processing and Control*, vol. 72, p. 103341, Feb. 2022, doi: 10.1016/j.bspc.2021.103341.

[14] P. M. Shafi, V. Bidve, H. Bhapkar, P. Dhotre, and V. B. P. Singh, "Leukemia detection system using convolutional neural networks by means of microscopic pictures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1616–1623, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1616-1623.

[15] B. Elsayed *et al.*, "Deep learning enhances acute lymphoblastic leukemia diagnosis and classification using bone marrow images," *Frontiers in Oncology*, vol. 13, p. 1330977, Dec. 2023, doi: 10.3389/fonc.2023.1330977.

[16] N. H. Shabrina, D. Gunawan, M. F. Ham, and A. S. Harahap, "Papillary Thyroid Cancer Histopathological Image Classification Using Pretrained ConvNeXt Tiny and Grad-CAM Interpretation," in *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Dec. 2023, pp. 1836–1842, doi: 10.1109/ITAIC58329.2023.10409019.

[17] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Computer Methods and Programs in Biomedicine*, vol. 226, p. 107161, Nov. 2022, doi: 10.1016/j.cmpb.2022.107161.

[18] P. J. G. Lisboa, S. Saralajew, A. Vellido, R. Fernández-Domenech, and T. Villmann, "The coming of age of interpretable and explainable machine learning models," *Neurocomputing*, vol. 535, pp. 25–39, May 2023, doi: 10.1016/j.neucom.2023.02.040.

[19] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A Review of Machine Learning Interpretability Methods," *Entropy*, vol. 23, no. 1, p. 18, Dec. 2020, doi: 10.3390/e23010018.

[20] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, USA, Jun. 2019, pp. 6105–6114.

[21] M. Tan and Q. V. Le, "EfficientNetV2: Smaller Models and Faster Training," in *Proceedings of the 38 th International Conference on Machine Learning*, Jul. 2021, pp. 10096–10106.

[22] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.293977.

[23] S. Mourya, S. Kant, P. Kumar, A. Gupta, and R. Gupta, "ALL Challenge dataset of ISBI 2019 (C-NMC 2019)," *The Cancer Imaging Archive*, 2019, doi: 10.7937/TCIA.2019.DC64I46R.

[24] "Keras Applications," Keras Applications, [Online]. Available: <https://keras.io/api/applications/>, (accessed: Jul. 03, 2024).

[25] D. P. Kingma, "Adam: A Method for Stochastic Optimization," *arXiv*, 2014, doi: 10.48550/arXiv.1412.6980.

[26] F. Mohr and J. N. van Rijn, "Learning Curves for Decision Making in Supervised Machine Learning -- A Survey," *arXiv*, 2022, doi: 10.48550/arXiv.2201.12150.

BIOGRAPHIES OF AUTHORS



Sion Theodorus Syaron Darmawan     received his bachelor's degree in computer engineering from Universitas Multimedia Nusantara. His interests include artificial intelligence and internet of things. He can be contacted at email: sion.theodorus@student.umn.ac.id.



Nabila Husna Shabrina     received her bachelor's degree in telecommunication engineering and master's degree in electrical engineering from Institut Teknologi Bandung. She is currently a lecturer in computer engineering at Universitas Multimedia Nusantara. Her research interests include computer vision and image processing. She can be contacted at email: nabila.husna@umn.ac.id.